

A two-queue model for optimising the value of information in energy-harvesting sensor networks

Kishor Patil^{a,*}, Koen De Turck^b, Dieter Fiems^a

^a*Ghent university, Department of Telecommunication and Information Processing, Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium*

^b*Laboratoire des Signaux et Systèmes, CentraleSupélec, 3, Rue Joliot-Curie, 91192 Gif-sur-Yvette, France*

Abstract

We consider a discrete-time queueing model with two coupled queues for studying the optimal transmission policy of an energy-harvesting sensor node. In particular, the sensor node under study operates energy neutral and harvests energy according to a Bernoulli process. Discretising energy into “energy chunks”, the battery is modelled as a first queue, whereas a second queue is introduced to hold the information at the sensor node. We consider a hybrid wireless sensor network in which a mobile sink is used to collect data. From the vantage point of the sensor node, this means that the sensor can only send when the sink is sufficiently close. When this is the case, the sensor decides whether to transmit its data or not depending on the amount of available energy and the value of the information (VoI). Focusing on the optimal transmission policy, we formulate the optimal control problem as a Markov Decision Process with a level-dependent block-triangular transition probability matrix. We find the optimal policy which maximises the mean VoI transmitted from the node in the long run. Finally, we investigate the structure of the optimal policy and the mean VoI collected from the node for different system parameters by means of some numerical experiments.

Keywords: Sensor network; Energy harvesting; Value of information; Markov decision process;

1. Introduction

Wireless sensor networks (WSNs) have attracted considerable research interest over the past few years because of their wide range of applications in health care, utilities, remote monitoring as well as in diverse industrial contexts [1]. WSNs mainly consist of three components: gateways, relay nodes, and sensors. Gateways act as an interface between wireless sensor nodes and the application platform. Relay nodes, sometimes referred as routers, are used to extend the coverage area of the sensor network. Finally, sensors can sense, measure and collect the information from the environment. The information sensed at the sensor node is analysed and the node can then decide whether to transmit the data or not. WSNs can be used in many applications that require close monitoring of the physical world which explains the wide range of areas in which they are applied.

Wireless sensor nodes are low power devices [2] equipped with a small battery and on-board memory. Currently, the primary power source for sensor nodes is a small on-board battery which limits the lifetime of sensors. Since they are often installed in hostile terrain, it is very expensive and difficult to replace the battery due to environmental and terrestrial challenges. To overcome this problem, more efficient energy consumption and power management are active areas of research and development. Some research focuses on energy conservation by controlling the communication system. As there is significant energy consumption when the sensor node is idle, switching off the sensor node can save the battery and thus extend the lifetime. In such strategies, the WSN node periodically wakes up to transmit the data and then sleeps by powering off to conserve the energy. WSN technology must efficiently manage the frequency of ON and OFF times of the node.

*Corresponding author

Email addresses: patil.kishor@ugent.be (Kishor Patil), koen.deturck@12s.centralesupelec.fr (Koen De Turck), dieter.fiems@ugent.be (Dieter Fiems)

In the context of sensor networks, a refinement of the familiar concept of Quality of Service (QoS) has been proposed, referred to as “Quality of Information” (QoI), which takes into account on the quality but also on the timeliness of the information [3, 4, 5]. In this paper, we make use of this new concept and keep track of the value of information that is currently stored in the sensor node.

To mitigate dependency on batteries, current research efforts aim to create sensors that harvest the necessary energy from the environment. These are known as energy-harvesting sensor nodes. They use ambient sources of energy present in the environment like solar, wind or heat and convert it into the electricity which can be used for transmission. However, the amount of energy harvested heavily depends on the environmental conditions which may lead to a fluctuating performance. Hence accurate models of WSNs should take into account the random nature of the harvesting process. Because of such heavy dependence on environmental factors, the role of power management becomes critical in energy-harvesting wireless sensor networks (EH-WSN).

In this paper, we aim to design an optimal transmission policy (which is essentially a set of decision rules describing the action to be taken in each state of the system) of a sensor node in a EH-WSN. We allow for the temporal non-availability of the sink by introducing so called transmission opportunities, which is a natural assumption when the sink in question is wirelessly connected or even mobile. We further assume that sensor nodes from time to time gather information to which we assign a value. Its value degrades over time as it is waiting for transmission in the memory of the sensor node. At each transmission opportunity, the sensor node decides whether to transmit the data or not depending on the amount of available energy and the value of information, both of which we assume to be discrete values. This paper investigates the structure of the optimal policy and the resulting optimal mean value of information transmitted from the sensor node. If the sensor node decides to transmit, it deletes the information from memory and hence the value of information present in the node resets to zero. From a queueing-theoretical point of view, such behaviour is also found in queues with disasters, which have been used amongst others in telecommunications and in particular in the context of satellites and in internet applications.

The remainder of this paper is organised as follows. In the next section, we discuss literature on Markov processes and Markov decision processes for energy harvesting sensor nodes. We then introduce the Markov decision process for age- and energy-dependent transmissions in sensor networks in section 3 and evaluate the optimal decision policies in section 4. Finally, we draw conclusions in Section 5.

2. Related Work

Optimal scheduling in energy harvesting sensor networks is an active research topic. We refer to Yick et al. [2] and Lu et al. [6] and the references therein for a discussion of key issues and developments in WSNs, addressing the different challenges including storage capacity, energy replenishment, network services, deployment etc. We restrict the discussion below to Markovian models of WSNs.

Such models have been studied by various authors, one of the first contributions being [7] where Susu et al. propose a finite birth-death Markov chain to model energy harvesting sensor nodes, a birth and death corresponding to energy harvesting and energy consumption respectively. The same energy harvesting process is considered by Jornet and Akyildiz [8], but the authors allow for more rapid energy consumption, a transmission corresponding to a deterministic jump downwards. Seyedi and Sikdar [9] propose a Markov model to study the trade-off between energy consumption and packet error probability. Allowing for time-correlation in the harvesting process, Ho et al. [10] and Lee et al. [11] assume a Markov modulated harvesting process and verify statistically that a Markov modulated process is indeed appropriate for describing solar energy harvesting. Most contributions study a single sensor node in isolation, a notable exception being [12] which studies a body sensor network with multiple harvesting sensors. Studying the complete network allows for assessing the probability of a node failing to detect an event owing to lack of energy, which is a key design consideration for body sensor sensors.

The models above, do not allow for buffering sensed data. Data buffering is important when sensors have limited connectivity. This is the case in hybrid sensor networks where data is collected by a mobile sink: data can only be offloaded when the mobile sink is within range. In contrast to Markovian models without data buffering, the Markov process now needs to capture both the state of the battery and the data buffer. De Cuyper et al. [13] consider such a two-dimensional Markov model and include a data buffer for temporary data storage if the SN is not connected or has insufficient energy to transmit. The resulting Markov process has a quasi-birth-death structure allowing for efficient numerical evaluation.

In addition to modelling harvesting sensor nodes, various authors also consider optimal control policies for transmissions. Such control is studied by Sharma et al. [14] for an energy harvesting sensor node that periodically transmits data. The authors discuss the existence of an optimal transmission policy but do not calculate it. Morsi et al. [15] calculate an on-off transmission policy either assuming a finite or infinite energy buffer. The same control problem is considered in [16, 17] in a deterministic setting: data arrivals and the amount of harvested energy are known in advance. Policies obtained from such model can be a benchmark but given the random nature of energy harvesting, one expects that the results cannot be attained in a random setting. Ozel et al. [18] consider the optimal scheduling for energy harvesting sensor node over a time varying channel. In an offline setting, the authors show the optimality of an adaptive directional water-filling algorithm for throughput maximisation whereas they rely on stochastic dynamic programming to solve for the optimal online policy.

Various authors introduce Markov decision processes (MDP) to study optimal transmission policies in wireless sensor networks, see the recent survey [19] and the references. Lei et al. [20] discuss energy-aware threshold transmission policies by modelling the energy state transitions by a continuous time Markov chain. Three different scenarios are considered based on recharging and/ or replacement of the battery model. Transmission policies depend on the condition of the channel state as well as on the value of the message. If the channel state is good then power required to transmit the data is less. Fernandez et al. [21] use the term “importance of message” for the value of the message and propose a stochastic approximation algorithm to find out the optimal transmission policy. The optimal policy is a threshold policy under reasonable assumptions on battery dynamics. Michelusi et al. [22] models the energy harvesting as a Markov process with two states, i.e., a good and a bad state. The authors adapt the transmission probabilities to the state of the battery, deriving the optimal policy using policy iteration. The optimal policy accounts for the data at sensor node and the amount of energy at the time of decision making. Some authors further focus on control policies when the energy harvesting distribution is unknown and the system can learn from past observations. [23, 24] discuss the optimal power control policies for EH-WSN by formulating the problem as a partially observable Markov decision process (POMDP). The authors consider an ARQ-based packet transmission protocol, their goal being to find the optimal transmit power for each packet transmission.

The present paper most closely relates to [13] in terms of modelling assumptions as we consider a model with both energy consumption and data buffering. However, this contribution also differs considerably from [13]. Foremost, we study non-additive data gathering dynamics, inspired by the concept of the QoI introduced above. Moreover, we formulate the associated MDP, while [13] studies the uncontrolled Markov model. Finally, the MDP problem at hand considerably differs from preceding MDP models for energy harvesting sensor nodes as the policies now account for both the state of the battery and the state of the data buffer.

3. Markov model

We focus on a single sensor node with energy harvesting capability in a sensor network where data is collected by a mobile sink as depicted in figure 1. The sensor has an on-board battery for storing harvested energy as well as the capability for storing sensed information. We assume that time is discrete. That is, time is divided into fixed length intervals (slots) and transmissions from the sensor node are synchronised with respect to slot boundaries.

The dynamics of the sensor node are governed by three independent stochastic processes. The first process describes how energy is harvested. Assuming that energy can be discretised into chunks of energy, let H_k denote the number of *energy chunks* or energy units harvested during the k th slot. The sequence $\{H_k, k \in \mathbb{N}\}$ constitutes a sequence of independent Bernoulli random variables. Let p_e denote the probability that a chunk of energy arrives during a slot. The second process describes how much information is sensed during a slot. Let D_k denote the amount of information sensed during the k th slot. We express the amount of information as a discrete number of *information units*. The process $\{D_k, k \in \mathbb{N}\}$ is a sequence of independent and identically distributed non-negative discrete random variables. The sensed information is at most M information units, $P[D_k > M] = 0$, and we denote the common probability mass function of the D_k 's by $d_m = P[D_k = m]$, $m = 0, \dots, M$. For further use, we also introduce the corresponding distribution function $\bar{d}_n = P[D_k \leq n] = \sum_{m \leq n} d_m$. The third process describes when the sensor node can pass on its information to the mobile sink. Let T_k be the indicator that there is a transmission opportunity at the k th slot boundary. The sequence $\{T_k, k \in \mathbb{N}\}$ constitutes a sequence of independent Bernoulli random variables, p_t denoting the probability that there is a transmission opportunity at a slot boundary.

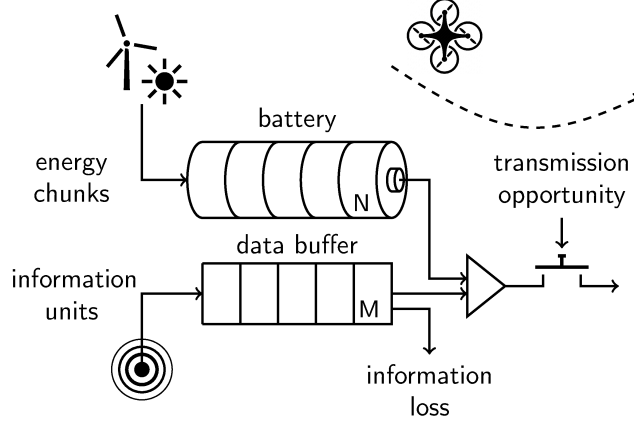


Figure 1: WSN Model

We can now describe the sensor node dynamics in terms of these random variables. Let B_k be the battery level (in terms of energy chunks) at the k th slot boundary. We assume that a single chunk of energy is required to transmit information to the mobile sink and that the on-board battery can store N chunks of energy at most. Energy chunks are lost when the battery capacity is exceeded. In view of these assumptions, we find that the battery level at the $(k + 1)$ st slot boundary can be expressed as,

$$B_{k+1} = \min(B_k + H_k - U_k, N).$$

Here U_k is the indicator that there is a transmission at the k th slot boundary. This random variable depends on the availability of the mobile sink T_k , on the battery level B_k as well as on the value of the information V_k as discussed below. In particular, transmissions are not possible when there is no energy, which means that $B_k = 0$ implies $U_k = 0$.

The evolution of the value of information V_k is the combination of an information loss process and an information replacement process. To capture the loss of information over time, we assume that the value of information decreases by one information unit per slot. In addition, we assume that the information at the sensor node is replaced by newly sensed information whenever the amount of newly sensed information units exceeds the value of information at the sensor node. Finally, whenever the information is passed on to the mobile sink, the information is deleted at the sensor node. In view of these assumptions, we can express the value of information at the $(k + 1)$ st slot boundary as follows,

$$V_{k+1} = \begin{cases} D_{k+1} & \text{for } U_k = 1, \\ \max(D_{k+1}, (V_k - 1)^+) & \text{for } U_k = 0. \end{cases} \quad (1)$$

Here U_k is again the indicator that there is a transmission at the k th slot boundary.

When there is no transmission opportunity, $T_k = 0$, there obviously is no transmission, hence $U_k = 0$. When there is a transmission opportunity, $T_k = 1$, the decision on transmitting depends on the current value of information, as well as on the battery level. As there is only energy consumption when there is a transmission, it can be beneficial to postpone transmissions when there is limited value of information at the sensor node. Therefore, we study the optimal transmission policy by modelling the sensor node as a Markov decision process (MDP).

3.1. MDP formulation

In order to define the MDP, we must specify its state space, action set, transition probabilities and rewards. We explain each of the components in detail below.

- (a) **State Space:** In view of the modelling assumptions above, the state of the sensor node at the k th slot boundary is described by the battery level B_k and the value of information V_k . These discrete random variables are positive and bounded by N (battery capacity) and M (maximum of the sensed value of information), respectively. In addition, as the decision to transmit or defer transmission depends on the availability of a transmission opportunity, we additionally track T_k as well.

Summarising, the state is described by the vector $(B_k, V_k, T_k) \in \mathcal{S} \triangleq \mathcal{B} \times \mathcal{V} \times \mathcal{T}$, where,

- $\mathcal{B} = \{0, 1, 2, \dots, N\}$ is the set of battery states,
- $\mathcal{V} = \{0, 1, 2, \dots, M\}$ is the set of the value of information states.
- $\mathcal{T} = \{0, 1\}$ is the set of the transmission opportunity states.

(b) **Action Set:** There are at most two possible actions: transmitting (action 1) or not transmitting (action 0). The sensor node cannot transmit in the absence of energy ($B_k = 0$) or in the absence of a transmission opportunity ($T_k = 0$), hence for $s \in \mathcal{S}_0 = \mathcal{B} \times \mathcal{V} \times \{0\} \cup \{0\} \times \mathcal{V} \times \mathcal{T}$, the only possible action is not to transmit: the action set is a singleton $\mathcal{A}_s = \{0\}$ for $s \in \mathcal{S}_0$. For $s \in \mathcal{S}_1 = \mathcal{S} \setminus \mathcal{S}_0$, there is a transmission opportunity as well as energy such that the action set includes both possible actions: $\mathcal{A}_s = \{0, 1\}$ for $s \in \mathcal{S}_1$.

For further use, let $\pi : \mathcal{S} \rightarrow \{0, 1\}$ denote a policy of the sensor node, that is a mapping from the state space to the action space $\{0, 1\}$. A policy π is admissible if $\pi(s) \in \mathcal{A}_s$ for all $s \in \mathcal{S}$. The policy relates to the indicator U_k of having a transmission at the k th slot boundary as,

$$U_k = \pi(B_k, V_k, T_k).$$

(c) **Transition Probabilities:** Let $q(s'|s) = q((i', j', k')|(i, j, k))$ represent the transition probability that state $s' = (i', j', k')$ is visited from state $s = (i, j, k)$ for $s \in \mathcal{S}_0$. We here suppress the action that is taken from the notation as for $s \in \mathcal{S}_0$ only action 0 is available. Remark first that the transition probabilities of the transmission opportunity process are given by

$$p^o(k'|k) = \begin{cases} 1 - p_t & \text{for } k' = 0, \\ p_t & \text{for } k' = 1, \end{cases}$$

which can also be written as $p^o(k'|k) = p_t^{k'}(1 - p_t)^{1-k'}$. Recalling that the entries of the state vector correspond to the battery level, the value of information and the indicator of the transmission opportunity respectively, we find,

$$q((i', j', k')|(i, j, k)) = \begin{cases} d_{j'} p_e p_t^{k'} (1 - p_t)^{1-k'} & \text{for } i' = i + 1 \text{ and } j' \geq j, \\ \bar{d}_{j-1} p_e p_t^{k'} (1 - p_t)^{1-k'} & \text{for } i' = i + 1 \text{ and } j' = j - 1, \\ d_{j'} (1 - p_e) p_t^{k'} (1 - p_t)^{1-k'} & \text{for } i' = i < N \text{ and } j' \geq j, \\ \bar{d}_{j-1} (1 - p_e) p_t^{k'} (1 - p_t)^{1-k'} & \text{for } i' = i < N \text{ and } j' = j - 1 \\ d_{j'} p_t^{k'} (1 - p_t)^{1-k'} & \text{for } i' = i = N \text{ and } j' \geq j, \\ \bar{d}_{j-1} p_t^{k'} (1 - p_t)^{1-k'} & \text{for } i' = i = N \text{ and } j' = j - 1, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

for $s \in \mathcal{S}_0$ and $s' \in \mathcal{S}$.

For $s \in \mathcal{S}_1$, let $p(s'|s, a) = p((i', j', k')|(i, j, k), a)$ represent the transition probability that state $s' = (i', j', k')$ is visited from state $s = (i, j, k)$ when action $a \in \{0, 1\}$ is chosen. For $a = 0$, the evolution of the state is similar to the evolution when there is no transmission opportunity. Hence, we have,

$$p((i', j', k')|(i, j, k), 0) = \begin{cases} d_{j'} p_e p_t^{k'} (1 - p_t)^{1-k'} & \text{for } i' = i + 1 \text{ and } j' \geq j, \\ \bar{d}_{j-1} p_e p_t^{k'} (1 - p_t)^{1-k'} & \text{for } i' = i + 1 \text{ and } j' = j - 1, \\ d_{j'} (1 - p_e) p_t^{k'} (1 - p_t)^{1-k'} & \text{for } i' = i < N \text{ and } j' \geq j, \\ \bar{d}_{j-1} (1 - p_e) p_t^{k'} (1 - p_t)^{1-k'} & \text{for } i' = i < N \text{ and } j' = j - 1 \\ d_{j'} p_t^{k'} (1 - p_t)^{1-k'} & \text{for } i' = i = N \text{ and } j' \geq j, \\ \bar{d}_{j-1} p_t^{k'} (1 - p_t)^{1-k'} & \text{for } i' = i = N \text{ and } j' = j - 1, \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

for $s \in \mathcal{S}_1$ and $s' \in \mathcal{S}$. Finally, for $a = 1$, we find,

$$p((i', j', k')|(i, j, k), 1) = \begin{cases} d_{j'} p_e p_t^{k'} (1 - p_t)^{1-k'} & \text{for } i' = i, \\ d_{j'} (1 - p_e) p_t^{k'} (1 - p_t)^{1-k'} & \text{for } i' = i - 1, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

for $s \in \mathcal{S}_1$ and $s' \in \mathcal{S}$.

- (d) **Reward:** We define the immediate reward in state $s \in S$ as the value of information that is transmitted to the mobile sink. That is, the reward in state $s \in \mathcal{S}_0$ is zero. For $s = (i, j, k) \in \mathcal{S}_1$ the reward equals j if $a = 1$ while there is no reward for $a = 0$. Let $R_\pi(s)$ be the immediate reward in state $s = (i, j, k)$ for policy π , we then have,

$$R_\pi(i, j, k) = \begin{cases} 0 & \text{for } \pi((i, j, k)) = 0 \\ j & \text{for } \pi((i, j, k)) = 1. \end{cases}$$

The immediate reward only depends on the state and the action.

3.2. Policy iteration

We focus on the infinite horizon control problem, assuming that a stationary policy is applied. Let $v_\pi(s)$ be the uniformly discounted value-to-go from state s under policy π ,

$$v_\pi(s) = v_\pi(i, j, k) = \mathbb{E} \left[\sum_{k=0}^{\infty} \alpha^k R_\pi(B_k, V_k, T_k) \middle| B_0 = i, V_0 = j, T_0 = k \right],$$

where α is a discounting factor. Further, let $v^*(s)$ be the optimal value-to-go from state s ,

$$v^*(s) = \max_{\pi} v_\pi(s),$$

where the maximum is taken over all admissible policies π . The optimal value-to-go and optimal policy can be found by solving the Bellman equations, which read,

$$v^*(s) = \sum_{s' \in S} \alpha q(s'|s) v^*(s')$$

for $s \in \mathcal{S}_0$, and,

$$v^*(s) = \max \left(j + \alpha \sum_{s' \in S} p(s'|s, 1) v^*(s'), \sum_{s' \in S} p(s'|s, a=0) \alpha v^*(s') \right)$$

for $s = (i, j, k) \in \mathcal{S}_1$.

To solve the Bellman equations, we rely on policy iteration [25]. To this end, starting from an initial policy π_0 , we iterative improve the policy. In each iteration $n = 0, 1, \dots$ we first solve the system of equations,

$$v_{\pi_n}(s) = \begin{cases} \sum_{s' \in S} q(s'|s) \alpha v_{\pi_n}(s') & \text{for } s \in \mathcal{S}_0, \\ j\pi_n(s) + \alpha \sum_{s' \in S} p(s'|s, \pi_n(s)) v_{\pi_n}(s') & \text{for } s \in \mathcal{S}_1. \end{cases} \quad (5)$$

We can then adjust the policy in the policy improvement step,

$$\pi_{n+1} = \arg \max_{\pi} \left(j\pi(s) + \alpha \sum_{s' \in S} p(s'|s, \pi(s)) v_{\pi_{n-1}}(s') \right).$$

In view of the above, we have $\pi_{n+1}(s) = 0$ for $s \in \mathcal{S}_0$ and

$$\pi_{n+1}(s) = 1 \left\{ j + \alpha \sum_{s' \in S} p(s'|s, 1) v_{\pi_{n-1}}(s') > \alpha \sum_{s' \in S} p(s'|s, 0) v_{\pi_{n-1}}(s') \right\} \quad (6)$$

for $s \in \mathcal{S}_1$ where $1\{\cdot\}$ is the indicator function.

3.3. Quasi birth death structure

The major computational effort in solving the MDP problem, is the evaluation of (5). Note however, that – regardless of the policy – the transitions of the battery level are skip-free in both directions: the battery level either remains the same or can only go up or down one level during a slot. This in turn means that the matrix corresponding to the system of equations (5) has a tridiagonal block structure or quasi birth death (QBD) structure, blocks grouping transitions from battery level i to battery level i' . As the battery level is skip free, the blocks are non-zero only for $i' \in \{i - 1, i, i + 1\}$ (hence the matrix is tridiagonal).

The block structure allows for efficiently solving the system of equations by linear level reduction [26]. To this end, let $\mathbf{v}_\pi(i, k)$ be the column vector with elements $v_\pi(i, j, k)$ for $j = 0, \dots, M - 1$. From equation (5), we then find that these vectors adhere,

$$\mathbf{v}_\pi(i, 0) = \alpha \bar{p}_t \bar{p}_e A \mathbf{v}_\pi(i, 0) + \alpha p_t \bar{p}_e A \mathbf{v}_\pi(i, 1) + \alpha \bar{p}_t p_e A \mathbf{v}_\pi(i + 1, 0) + \alpha p_t p_e A \mathbf{v}_\pi(i + 1, 1) \quad (7)$$

$$\mathbf{v}_\pi(N, 0) = \alpha \bar{p}_t A \mathbf{v}_\pi(N, 0) + \alpha p_t A \mathbf{v}_\pi(N, 1) \quad (8)$$

for $i = 0, \dots, N - 1$ and,

$$\mathbf{v}_\pi(0, 1) = \alpha \bar{p}_t \bar{p}_e A \mathbf{v}_\pi(0, 0) + \alpha p_t \bar{p}_e A \mathbf{v}_\pi(0, 1) + \alpha \bar{p}_t p_e A \mathbf{v}_\pi(1, 0) + \alpha p_t p_e A \mathbf{v}_\pi(1, 1) \quad (9)$$

$$\begin{aligned} \mathbf{v}_\pi(i, 1) = & \Pi(i)\theta + \alpha \Pi(i) \bar{p}_t \bar{p}_e B \mathbf{v}_\pi(i - 1, 0) + \alpha \Pi(i) p_t \bar{p}_e B \mathbf{v}_\pi(i - 1, 1) + \alpha \Pi(i) \bar{p}_t p_e B \mathbf{v}_\pi(i, 0) + \alpha \Pi(i) p_t p_e B \mathbf{v}_\pi(i, 1) \\ & + \alpha \bar{\Pi}(i) \bar{p}_e \bar{p}_t A \mathbf{v}_\pi(i, 0) + \alpha \bar{\Pi}(i) \bar{p}_e p_t A \mathbf{v}_\pi(i, 1) + \alpha \bar{\Pi}(i) p_e \bar{p}_t A \mathbf{v}_\pi(i + 1, 0) + \alpha \bar{\Pi}(i) p_e p_t A \mathbf{v}_\pi(i + 1, 1) \end{aligned} \quad (10)$$

$$\begin{aligned} \mathbf{v}_\pi(N, 1) = & \Pi(i)\theta + \alpha \Pi(i) \bar{p}_t \bar{p}_e B \mathbf{v}_\pi(N - 1, 0) + \alpha \Pi(i) p_t \bar{p}_e B \mathbf{v}_\pi(N - 1, 1) + \alpha \Pi(i) \bar{p}_t p_e B \mathbf{v}_\pi(N, 0) \\ & + \alpha \Pi(i) p_t p_e B \mathbf{v}_\pi(N, 1) + \alpha \bar{\Pi}(i) \bar{p}_t A \mathbf{v}_\pi(N, 0) + \alpha \bar{\Pi}(i) p_t A \mathbf{v}_\pi(N, 1) \end{aligned} \quad (11)$$

for $i = 1, \dots, N - 1$ with $\theta = [0, 1, 2, \dots, M]'$ a column vector with $M + 1$ entries, with $\Pi(i)$ the $(M + 1) \times (M + 1)$ diagonal matrix with diagonal entries $\pi(i, j, 1)$ ($j = 0, \dots, M - 1$), with $\bar{\Pi}(i)$ the $(M + 1) \times (M + 1)$ diagonal matrix with diagonal entries $1 - \pi(i, j, 1)$ ($j = 0, \dots, M - 1$), and where the $(M + 1) \times (M + 1)$ matrices A and B are defined as follows,

$$A = \begin{bmatrix} d_0 & d_1 & d_2 & d_3 & \cdots & d_M \\ \bar{d}_0 & d_1 & d_2 & d_3 & \cdots & d_M \\ 0 & \bar{d}_1 & d_2 & d_3 & \cdots & d_M \\ 0 & 0 & \bar{d}_2 & d_3 & \cdots & d_M \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & d_M \end{bmatrix}, \quad B = \begin{bmatrix} d_0 & d_1 & d_2 & d_3 & \cdots & d_M \\ d_0 & d_1 & d_2 & d_3 & \cdots & d_M \\ d_0 & d_1 & d_2 & d_3 & \cdots & d_M \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ d_0 & d_1 & d_2 & d_3 & \cdots & d_M \end{bmatrix}.$$

We further introduce the column vectors $\mathbf{v}_\pi(i) = [\mathbf{v}_\pi(i, 0)', \mathbf{v}_\pi(i, 1)']'$. From equations (7) and (9), we find,

$$\mathbf{v}_\pi(0) = \alpha \bar{p}_e \begin{bmatrix} \bar{p}_t A & p_t A \\ \bar{p}_t A & p_t A \end{bmatrix} \mathbf{v}_\pi(0) + \alpha p_e \begin{bmatrix} \bar{p}_t A & p_t A \\ \bar{p}_t A & p_t A \end{bmatrix} \mathbf{v}_\pi(1). \quad (12)$$

Analogously, equations (7) and (10) yield,

$$\begin{aligned} \mathbf{v}_\pi(i) = & \begin{bmatrix} 0 \\ \Pi(i)\theta \end{bmatrix} + \alpha \bar{p}_e \begin{bmatrix} 0 & 0 \\ \bar{p}_t \Pi(i) B & p_t \Pi(i) B \end{bmatrix} \mathbf{v}_\pi(i - 1) - \alpha \bar{p}_e \begin{bmatrix} 0 & 0 \\ \Pi(i) \bar{p}_t A & \Pi(i) p_t A \end{bmatrix} \mathbf{v}_\pi(i) \\ & + \alpha p_e \begin{bmatrix} 0 & 0 \\ \Pi(i) \bar{p}_t B & \Pi(i) p_t B \end{bmatrix} \mathbf{v}_\pi(i) + \alpha \bar{p}_e \begin{bmatrix} \bar{p}_t A & p_t A \\ \bar{p}_t A & p_t A \end{bmatrix} \mathbf{v}_\pi(i) + \alpha p_e \begin{bmatrix} \bar{p}_t A & p_t A \\ \bar{\Pi}(i) \bar{p}_t A & \bar{\Pi}(i) p_t A \end{bmatrix} \mathbf{v}_\pi(i + 1), \end{aligned} \quad (13)$$

whereas equation (8) and (11) lead to,

$$\begin{aligned} \mathbf{v}_\pi(N) = & \begin{bmatrix} 0 \\ \Pi(N)\theta \end{bmatrix} + \alpha \bar{p}_e \begin{bmatrix} 0 & 0 \\ \bar{p}_t \Pi(N) B & p_t \Pi(N) B \end{bmatrix} \mathbf{v}_\pi(N - 1) \\ & + \alpha p_e \begin{bmatrix} 0 & 0 \\ \Pi(N) \bar{p}_t B & \Pi(N) p_t B \end{bmatrix} \mathbf{v}_\pi(N) + \alpha \begin{bmatrix} \bar{p}_t A & p_t A \\ \bar{\Pi}(N) \bar{p}_t A & \bar{\Pi}(N) p_t A \end{bmatrix} \mathbf{v}_\pi(N). \end{aligned} \quad (14)$$

The former set of matrices confirm the QBD structure of the system of equations, as $\mathbf{v}_\pi(i)$ is expressed in terms of $\mathbf{v}_\pi(i - 1)$, $\mathbf{v}_\pi(i)$ and $\mathbf{v}_\pi(i + 1)$ for all i .

3.4. Linear level reduction

As mentioned above, the structural property of the system of equations allows for solving the system of equations efficiently by linear level reduction [26]. Let the $2(M+1) \times 2(M+1)$ matrices F_0 and G_0 , and the $2(M+1)$ column vector h_0 be defined as,

$$F_0 = \left(I - \alpha \bar{p}_e \begin{bmatrix} \bar{p}_t A & p_t A \\ \bar{p}_t A & p_t A \end{bmatrix} \right)^{-1}, \quad G_0 = F_0 \alpha p_e \begin{bmatrix} \bar{p}_t A & p_t A \\ \bar{p}_t A & p_t A \end{bmatrix}, \quad h_0 = [0, 0, \dots, 0]'. \quad (15)$$

Here I is the $2(M+1) \times 2(M+1)$ identity matrix. Hence, we have $\mathbf{v}_\pi(0) = G_0 \mathbf{v}_\pi(1) + h_0$. For $0 < i < N$, we again have $\mathbf{v}_\pi(i) = G_i \mathbf{v}_\pi(i+1) + h_i$ where

$$G_i = \alpha p_e F_i \begin{bmatrix} \bar{p}_t A & p_t A \\ \bar{\Pi}(i) \bar{p}_t A & \bar{\Pi}(i) p_t A \end{bmatrix}, \quad h_i = F_i \left(\begin{bmatrix} 0 \\ \Pi(i) \theta \end{bmatrix} + \alpha \bar{p}_e \begin{bmatrix} 0 & 0 \\ \bar{p}_t \Pi(i) B & p_t \Pi(i) B \end{bmatrix} h_{i-1} \right) \quad (16)$$

with,

$$F_i = \left(I - \alpha p_e \begin{bmatrix} 0 & 0 \\ \Pi(i) \bar{p}_t B & \Pi(i) p_t B \end{bmatrix} - \alpha \bar{p}_e \begin{bmatrix} \bar{p}_t A & p_t A \\ \bar{p}_t A & p_t A \end{bmatrix} + \alpha \bar{p}_e \begin{bmatrix} 0 & 0 \\ \Pi(i) \bar{p}_t A & \Pi(i) p_t A \end{bmatrix} - \alpha \bar{p}_e \begin{bmatrix} 0 & 0 \\ \bar{p}_t \Pi(i) B & p_t \Pi(i) B \end{bmatrix} G_{i-1} \right)^{-1}.$$

Finally, for battery level N , we have,

$$\mathbf{v}_\pi(N) = \left(I - \alpha \bar{p}_e \begin{bmatrix} 0 & 0 \\ \bar{p}_t \Pi(N) B & p_t \Pi(N) B \end{bmatrix} G_{N-1} - \alpha p_e \begin{bmatrix} 0 & 0 \\ \Pi(N) \bar{p}_t B & \Pi(N) p_t B \end{bmatrix} - \alpha \begin{bmatrix} \bar{p}_t A & p_t A \\ \bar{\Pi}(N) \bar{p}_t A & \bar{\Pi}(N) p_t A \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} 0 \\ \Pi(N) \theta \end{bmatrix} + \alpha \bar{p}_e \begin{bmatrix} 0 & 0 \\ \bar{p}_t \Pi(N) B & p_t \Pi(N) B \end{bmatrix} h_{N-1} \right). \quad (17)$$

We find $\mathbf{v}_\pi(N)$ by this last equation, and can then recursively calculate all $\mathbf{v}_\pi(i)$ as we have $\mathbf{v}_\pi(i) = G_i \mathbf{v}_\pi(i+1) + h_i$ for $i < N$.

3.5. Summary and computational complexity

Summarising, we obtain the optimal policy π as follows,

1. Start with an initial policy π_0 . As we have $\pi_0(s) = 0$ for $s \in \mathcal{S}_0$, we only need to choose an action for all $s \in \mathcal{S}_1$. Set $n = 0$.
2. Given $\pi = \pi_n$, we calculate $\mathbf{v}_\pi(i)$ for $i = 0, \dots, N$ as follows:
 - (a) Calculate F_0 , G_0 and h_0 in accordance with equation (15).
 - (b) Calculate F_i , G_i and h_i for $i = 1, \dots, N-1$ in accordance with equation (16).
 - (c) Calculate $\mathbf{v}_\pi(N)$ using equation (17).
 - (d) Recursively calculate $\mathbf{v}_\pi(i)$ for $i = N-1, \dots, 0$ using $\mathbf{v}_\pi(i) = G_i \mathbf{v}_\pi(i+1) + h_i$.
Finally, set $\mathbf{v}_{\pi_n}(i) = \mathbf{v}_\pi(i)$ for $i = 0, \dots, N$.
3. We can now update the policy. For each $s \in \mathcal{S}_0$, set $\pi_{n+1}(s) = 0$. For each $s \in \mathcal{S}_1$ calculate $\pi_{n+1}(s)$ in accordance with equation (6).
4. If $\pi_{n+1} = \pi_n$, return π_n , if not, set $n \leftarrow n + 1$ and return to 2.

As policy iteration assures that the policy improves in every step, the number of iterations is bounded and the algorithm is guaranteed to converge in a finite number of iterations [27]. Each iteration requires the solution of the system of linear equations which has complexity $O(NM^3)$, see [26] for a general discussion on QBDs or [28] for a discussion on QBDs in the context of solving MDPs.

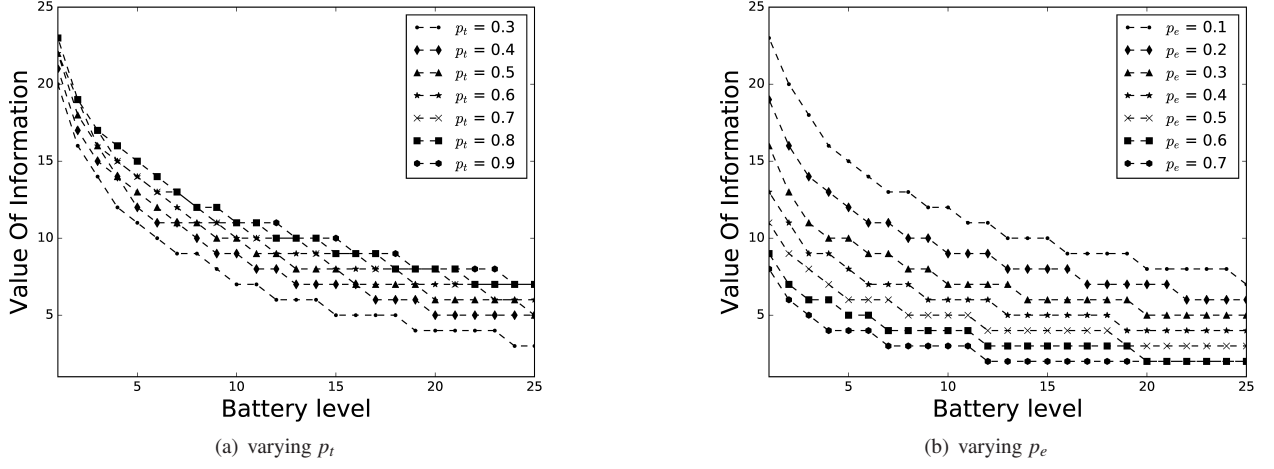


Figure 2: Optimal policy (a) for varying Transmission opportunity p_t (b) Varying harvesting energy probability p_e as indicated.

4. Numerical examples and discussion

We now investigate the structure of the optimal policy with respect to the different system parameters. To study the optimal policy for different transmission opportunity probabilities p_t , we fix the maximum capacity of the battery to $N = 100$ energy chunks and the VoI is at most $M = 100$ information units, the distribution being,

$$d_i = p(1 - p)^i.$$

for $0 < i \leq M$ with $p = 0.1$ and $d_0 = 1 - \sum_{k=1}^M d_k$.

The optimal policy obtained is a deterministic threshold policy. As the system has only one action available when there is no transmission opportunity, we analyse the policy for the reduced state space \mathcal{S}_1 . The optimal threshold is two dimensional, meaning that for a fixed VoI we always transmit beyond a certain battery level and that for a fixed battery level we always transmit beyond a certain number of information units.

Figures 2(a) and 2(b) depict the threshold for the VoI as a function of the battery state for a fixed discounting factor $\alpha = 0.9$. Figure 2(a) fixes the probability of an energy arrival $p_e = 0.1$ and shows the threshold for various transmission opportunity probabilities as indicated. Figure 2(b) fixes the probability of a transmission opportunity $p_t = 0.9$ and shows the threshold for different energy arrival probabilities as indicated.

It can be seen from both figures that the threshold decreases for increasing levels of the battery. This is not unexpected. When only a little energy is available, the decision to transmit affects future transmissions more if there is but a little energy such that one only transmits if there is a lot of information. If more energy available, the effect on future transmissions is smaller. Moreover, an additional increase will then hardly influence the threshold. Figure 2(a) further reveals that the threshold increases for increasing values of p_t . If there are many transmission opportunities, the chance to send more data units later on increases as the next transmission opportunity is not far away in time. It is also observed that the threshold for higher transmission opportunity probabilities can be equal to the threshold at lower transmission opportunity probabilities but cannot be less. In addition, figure 2(b) shows that boosting the energy harvesting capability yields lower thresholds. Higher p_e implies that availability of energy at future transmission opportunities is more likely so that one can send even when there is less data. If the battery has enough energy, a further increase in harvesting capability cannot improve the optimal reward and thus the threshold does not vary significantly.

Figure 3(a) shows the VoI threshold versus the battery level for different discounting factors as indicated. We fix the transmission opportunity probability $p_t = 0.9$ and energy arrival probability $p_e = 0.1$, while all the other parameters are kept the same. A higher discounting factor means that more importance is given to the future rewards. That is, when the discounting factor is high, the system is more likely to conserve the energy for messages which may arrive in future with higher VoI which implies that the threshold for sending the VoI is higher as well. Further notice

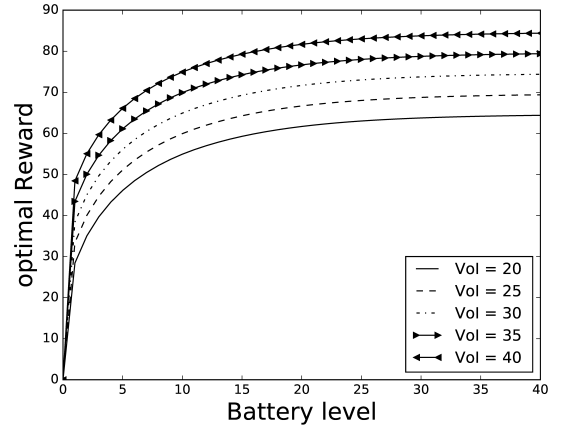
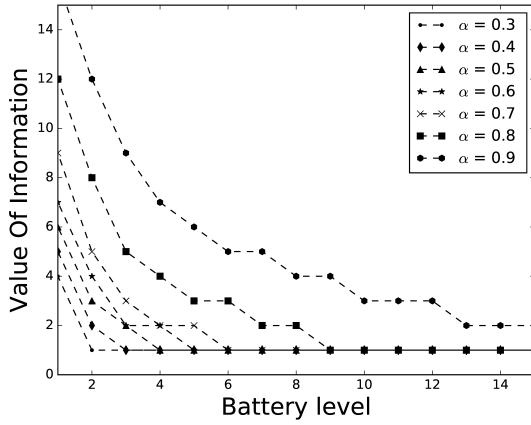
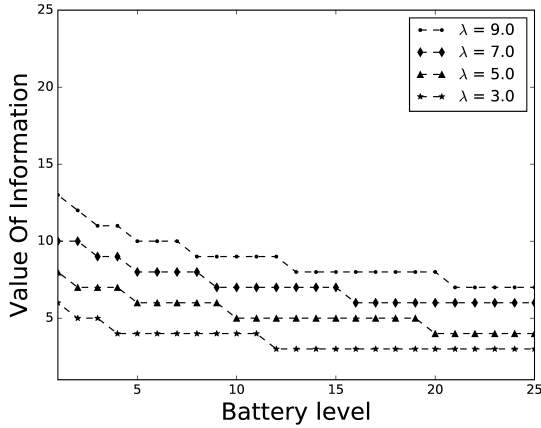
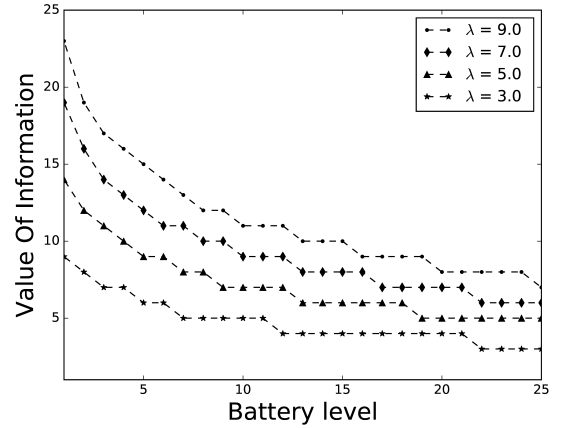


Figure 3: (a) Optimal policy for a varying discounting factor and (b) reward at the optimal policy as indicated.



(a) Poisson



(b) Geometric

Figure 4: Optimal policy for (a) Poisson (b) geometric VoI distribution for different values of mean (λ) of data arrival as indicated.

that the threshold for low discount factors can equal the threshold for a higher discount factor but cannot exceed that threshold. Figure 3(b) shows the reward $v^*(i, j, 1)$ for the optimal policy versus the battery level i for different values of the value of information j as indicated. We retain the parameters of figure 3(a), and additionally fix the discount factor to $\alpha = 0.9$. We observe that for a fixed VoI, the optimal reward is a non-decreasing function of the battery level. Results not depicted here show that if we fix the battery level, the optimal expected reward is non-decreasing for different VoI as well. With some additional effort, this can be shown theoretically using the approach given by Kashef et al.[29].

We now focus on the effects of the distribution of the value of Information on the optimal policy. To this end, figures 4(a) and 4(b) depict the threshold for (truncated) Poisson and (truncated) geometric distributed VoI respectively. Different values for the mean number of information units (λ) are assumed as indicated. We assume that the battery can store up to $N = 50$ chunks and that the VoI is bounded by $M = 50$. Moreover, the discounting factor is equal to $\alpha = 0.9$, the transmission opportunity probability is assumed to be $p_t = 0.9$ and the energy arrival probability is fixed to $p_e = 0.1$. Comparing figures 4(a) and 4(b) reveals that the distribution of the VoI affects the threshold policy significantly.

Optimal data collection

To conclude we study the optimal collection probability p_t , assuming that (i) the sensor node applies the optimal policy and (ii) there is a cost related to data collection.

To this end, we calculate the mean value of information collected per time slot as a function of p_t . For every value of p_t , we first find the optimal policy in accordance with section 3. Given this policy, we obtain the Markov chain for the optimally controlled sensor node. We have the transition probabilities,

$$\tau(s'|s) = \begin{cases} \sum_{s' \in \mathcal{S}} q(s'|s) & \text{for } s \in \mathcal{S}_0, \\ \sum_{s' \in \mathcal{S}} p(s'|s, \pi^*(s)) & \text{for } s \in \mathcal{S}_1, \end{cases} \quad (18)$$

with $q(s'|s)$ and $p(s'|s)$ as defined in equations (2) - (4). Let $\gamma(s)$ be the stationary probability of being in state s . That is, $\gamma(s)$ is the normalised solution of the system of equations,

$$\gamma(s') = \sum_{s \in \mathcal{S}} \tau(s'|s) \gamma(s),$$

for $s' \in \mathcal{S}$. Note that the system of equations above is again a QBD. Hence, we can again rely on linear level reduction to calculate these probabilities in $O(NM^3)$, see Latouche and Ramaswami [26].

Once we have found $\gamma(s)$, the mean reward per time slot can be expressed as,

$$\bar{V} = \sum_{s=(i,j,k) \in \mathcal{S}} \gamma(s) j \pi(s).$$

We now investigate the optimal data collection probability for the sensor node at hand, similarly as in [30] for the uncontrolled sensor node. We assume that there is a cost c associated to data collection such that the average value after collection equals,

$$\bar{V}_p = -cp_t + \bar{V}. \quad (19)$$

Note that we hereby assume that the cost can be expressed in terms of information units.

We can now study the impact of transmission opportunity probability on the average VoI collected from a sensor node operating under its optimal policy. We fix the battery capacity to 50 chunks and the range of VoI is 1 to 50 i.e., $N = 50, M = 50$.

Figure 5(a) fixes the energy arrival probability to 0.4 and plots the \bar{V}_p for different discounting factors whereas figure 5(b) fixes the discounting factor to 0.4 and shows the average VoI collected from node for different energy arrival probabilities. The cost of collection is assumed to be 2 i.e., $c = 2$. One clearly observes that \bar{V}_p increases for increasing transmission opportunities and then decreases again. This can be explained by the fact that by increasing p_t one first increases the amount of data one can collect. However, once p_t is sufficiently high, hardly any additional value can be collected by further increasing p_t . In contrast the collection cost does increase such that the overall value decreases.

The small jumps in the figures are present due to a change of the optimal policy at particular transmission opportunity probabilities. As a consequence, the curve representing \bar{V}_p is not concave due to these jumps and there may exist multiple optimum transmission opportunity probabilities for which \bar{V}_p attains its maximum value. It is also observed that the optimal transmission opportunity is more sensitive to changes in p_e than to changes of the discounting factor.

5. Conclusions

In this paper, we have proposed a two-queue Markov model for a wireless sensor node. The introduction of “Value of Information” and its non-additive data gathering process plays an important role in decision making. We formulated the problem as Markov decision process and found the exact solution by policy iteration. Numerical results show that the optimal transmission policy is a threshold policy. Further observations show that the threshold for the value of the information is most sensitive to the battery level when there is but few energy: the node transmits more selectively when there is less energy. We also studied the behaviour of the optimal reward and mean value of information collected from a node under the optimal policy, where we found the pervasive structural property that the optimal reward is non-decreasing in terms of the battery level as well as the VoI.

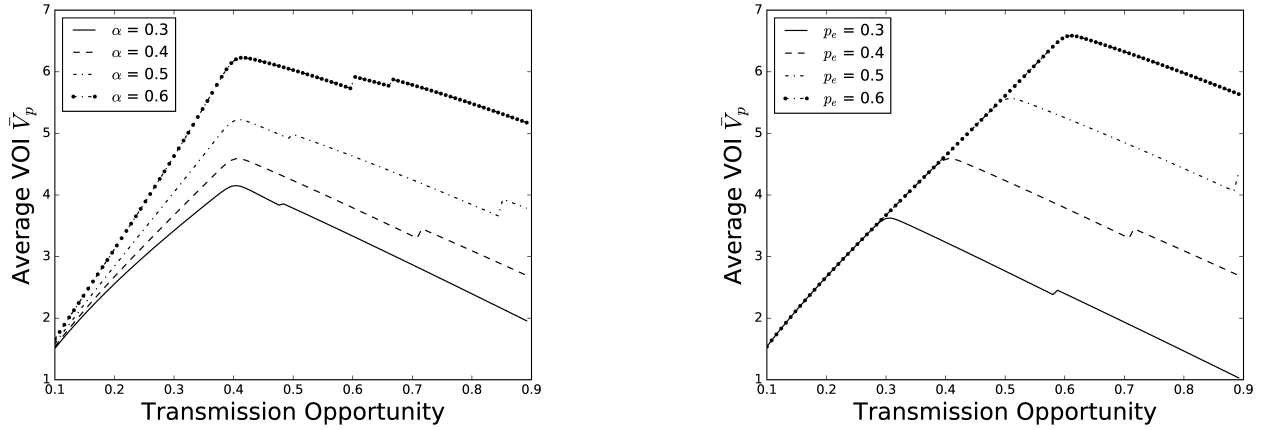


Figure 5: Value of information collected from node for (a) varying discounting factor and (b) varying energy harvesting probability

References

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, Wireless sensor networks: a survey, *Computer Networks* 38 (2002) 393–422.
- [2] J. Yick, B. Mukherjee, D. Ghosal, Wireless sensor network survey, *Computer Networks* 52 (12) (2008) 2292–2330.
- [3] V. Sachidananda, A. Khelil, N. Suri, Quality of information in wireless sensor networks: A survey, In: *Proc. of The International Conference on Information Quality (ICIQ)*, 2010.
- [4] C. Bisdikian, L. M. Kaplan, M. B. Srivastava, On the quality and value of information in sensor networks, *ACM Transactions on Sensor Networks* 9 (4) (2013) 48.
- [5] E. C.-H. Ngai, P. Gunningberg, Quality-of-information-aware data collection for mobile sensor networks, *Pervasive and Mobile Computing* 11 (2014) 203–215.
- [6] X. Lu, P. Wang, D. Niyato, D. Kim, Z. Han, Wireless networks with rf energy harvesting: A contemporary survey, *IEEE Communications Surveys & Tutorials* 17 (2) (2015) 757–789.
- [7] A. E. Susu, A. Acquaviva, D. Atienza, G. De Micheli, Stochastic modeling and analysis for environmentally powered wireless sensor nodes, in: *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks and Workshops*, 2008. *WiOPT 2008*. 6th International Symposium on, IEEE, 2008, pp. 125–134.
- [8] J. M. Jornet, I. F. Akyildiz, Joint energy harvesting and communication analysis for perpetual wireless nanosensor networks in the terahertz band, *IEEE Trans. Nanotechnol.* 11 (3) (2012) 570–580.
- [9] A. Seyed, B. Sikdar, Energy efficient transmission strategies for Body Sensor Networks with energy harvesting, 2008 42nd Annu. Conf. Inf. Sci. Syst. 58 (7) (2008) 2116–2126.
- [10] C. K. Ho, P. D. Khoa, P. C. Ming, Markovian models for harvested energy in wireless communications, 12th IEEE Int. Conf. Commun. Syst. 2010, ICCS 2010 (2010) 311–315.
- [11] P. Lee, Z. A. Eu, M. Han, H.-P. Tan, Empirical modeling of a solar-powered energy harvesting wireless sensor node for time-slotted operation, in: *2011 IEEE Wireless Communications and Networking Conference*, IEEE, 2011, pp. 179–184.
- [12] J. Ventura, K. Chowdhury, Markov modeling of energy harvesting body sensor networks, in: *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, IEEE, 2011, pp. 2168–2172.
- [13] E. De Cuyper, K. De Turck, D. Fiems, Stochastic modelling of energy harvesting for low power sensor nodes, in: *7th International Conference on Queueing Theory and Network Applications*, Proceedings, 2012, p. 6.
- [14] V. Sharma, U. Mukherji, V. Joseph, S. Gupta, Optimal energy management policies for energy harvesting sensor nodes, *IEEE Transactions on Wireless Communications* 9 (4) (2010) 1326–1336.
- [15] R. Morsi, D. S. Michalopoulos, R. Schober, On-off transmission policy for wireless powered communication with energy storage, in: *2014 48th Asilomar Conference on Signals, Systems and Computers*, IEEE, 2014, pp. 1676–1682.
- [16] K. Tutuncuoglu, A. Yener, Optimum transmission policies for battery limited energy harvesting nodes, *IEEE Transactions on Wireless Communications* 11 (3) (2012) 1180–1189.
- [17] F. M. Ozelik, G. Uctu, E. Uysal-Biyikoglu, Optimal packet scheduling on an energy harvesting fading channel, *CoRR*.
- [18] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, A. Yener, Transmission with energy harvesting nodes in fading wireless channels: Optimal policies, *IEEE J. Sel. Areas Commun.* 29 (8) (2011) 1732–1743.
- [19] M. Abu Alsheikh, D. T. Hoang, D. Niyato, H.-P. Tan, S. Lin, Markov Decision Processes With Applications in Wireless Sensor Networks: A Survey, *IEEE Commun. Surv. Tutorials* 17 (3) (2015) 1239–1267.
- [20] J. Lei, R. Yates, L. Greenstein, A generic model for optimizing single-hop transmission policy of replenishable sensors, *IEEE Trans. Wirel. Commun.* 8 (2) (2009) 547–551.
- [21] J. Fernandez-Bes, J. Cid-Sueiro, A. G. Marques, An MDP Model for Censoring in Harvesting Sensors: Optimal and Approximated Solutions, *IEEE J. Sel. Areas Commun.* 33 (8) (2015) 1717–1729.

- [22] N. Michelusi, K. Stamatiou, M. Zorzi, Transmission policies for energy harvesting sensors with time-correlated energy supply, *IEEE Transactions on Communications* 61 (7) (2013) 2988–3001.
- [23] A. Aprem, C. R. Murthy, N. B. Mehta, Transmit Power Control Policies for Energy Harvesting Sensors With Retransmissions, *IEEE J. Sel. Top. Signal Process.* 7 (5) (2013) 895–906.
- [24] A. Aprem, C. R. Murthy, N. B. Mehta, Transmit power control with ARQ in energy harvesting sensors: A decision-theoretic approach, in: 2012 IEEE Glob. Commun. Conf., IEEE, 2012, pp. 3388–3393.
- [25] D. P. Bertsekas, S. E. Shreve, *Stochastic Optimal Control: The Discrete-Time Case*, Academic Press, 1978.
- [26] G. Latouche, V. Ramaswami, *Introduction to Matrix Analytic Methods in Stochastic Modeling*, ASA-SIAM Series on Statistics and Applied Probability, Society for Industrial and Applied Mathematics, 1999.
- [27] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley & Sons, Inc. Hoboken, New Jersey, 2005.
- [28] L. B. White, A new policy evaluation algorithm for markov decision processes with quasi birth-death structure, *Stochastic Models* 21 (2-3) (2005) 785–797.
- [29] M. Kashef, A. Ephremides, Optimal scheduling for energy harvesting sources on time varying wireless channels, 2011 49th Annu. Allert. Conf. Commun. Control. Comput. Allert. 2011 14 (2) (2011) 712–718.
- [30] K. Patil, K. De Turck, D. Fiems, Optimal data collection in hybrid energy-harvesting sensor networks, in: *International Conference on Analytical and Stochastic Modeling Techniques and Applications*, Springer LNCS, 2016, pp. 239–252.